B11

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : G06K 9/00 — A1

(11) International Publication Number: WO 96/36007

(43) International Publication Date: 14 November 1996 (14.11.96)

(21) International Application Number: PCT/US96/06542

(22) International Filing Date: 10 May 1996 (10.05.96)

(30) Priority Data:
08/435,439 10 May 1995 (10.05.95) US

(71) Applicant: MAGIC CIRCLE MEDIA, INC. [US/US]; 2 North First Street, San Jose, CA 95113 (US).

(72) Inventors: ISADORE-BARRECA, Anthony, J.; 2312 Tiffin Road, Oakland, CA 94602 (US). EMBERSON, David, R.; 300 Moore Creek Road, Santa Cruz, CA 95060 (US). HIMELSTEIN, Mark, I.; 12256 Candy Court, Saratoga, CA 95070 (US). MEYER, Donald, E.; 10929 Northshore Square, Cupertino, CA 95014 (US).

(74) Agents: DULIN, Jacques et al.; Oblon, Spivak, McClelland, Maier & Neustadt, P.C., Crystal Square Five, 4th floor, 1755 Jefferson Davis Highway, Arlington, VA 22202 (US).

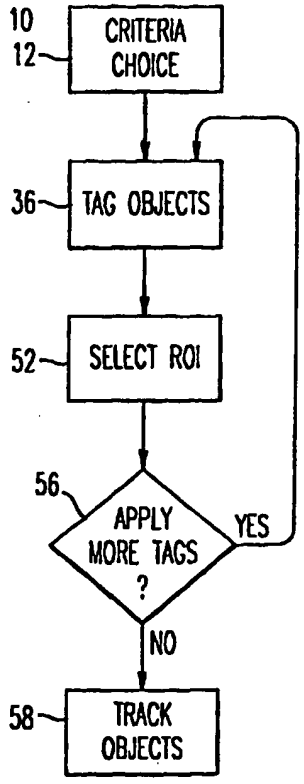
(81) Designated States: AL, AM, AT, AU, AZ, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IS, JP, KE, KG, KP, KR, KZ, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, UZ, VN, ARIPO patent (KE, LS, MW, SD, SZ, UG), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: OBJECT IDENTIFICATION IN A MOVING VIDEO IMAGE

(57) Abstract

A moving object identification method (10) for identifying and tracing an object (20) within a video image (14) such that the object (20) can act as a hot spot (30) as for an interactive computer/user interface (70). A plurality of tags (28) define the edges (26) of the object (20) and a plurality of hot spot borders (32) define the hot spot (30) such that the hot spot (30) and the object (20) generally coincide. A physical tag (28b) is optionally used to position the tags (28). Sensitivity to disapperance of the edges (26) is adjustable according to the relative size of a subtag (68) to the corresponding tag (28).



FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	**GB**	United Kingdom	**MW**	Malawi
AT	Austria	**GE**	Georgia	**MX**	Mexico
AU	Australia	**GN**	Guinea	**NE**	Niger
BB	Barbados	**GR**	Greece	**NL**	Netherlands
BE	Belgium	**HU**	Hungary	**NO**	Norway
BF	Burkina Faso	**IE**	Ireland	**NZ**	New Zealand
BG	Bulgaria	**IT**	Italy	**PL**	Poland
BJ	Benin	**JP**	Japan	**PT**	Portugal
BR	Brazil	**KE**	Kenya	**RO**	Romania
BY	Belarus	**KG**	Kyrgystan	**RU**	Russian Federation
CA	Canada	**KP**	Democratic People's Republic of Korea	**SD**	Sudan
CF	Central African Republic			**SE**	Sweden
CG	Congo	**KR**	Republic of Korea	**SG**	Singapore
CH	Switzerland	**KZ**	Kazakhstan	**SI**	Slovenia
CI	Côte d'Ivoire	**LI**	Liechtenstein	**SK**	Slovakia
CM	Cameroon	**LK**	Sri Lanka	**SN**	Senegal
CN	China	**LR**	Liberia	**SZ**	Swaziland
CS	Czechoslovakia	**LT**	Lithuania	**TD**	Chad
CZ	Czech Republic	**LU**	Luxembourg	**TG**	Togo
DE	Germany	**LV**	Latvia	**TJ**	Tajikistan
DK	Denmark	**MC**	Monaco	**TT**	Trinidad and Tobago
EE	Estonia	**MD**	Republic of Moldova	**UA**	Ukraine
ES	Spain	**MG**	Madagascar	**UG**	Uganda
FI	Finland	**ML**	Mali	**US**	United States of America
FR	France	**MN**	Mongolia	**UZ**	Uzbekistan
GA	Gabon	**MR**	Mauritania	**VN**	Viet Nam

OBJECT IDENTIFICATION IN A MOVING VIDEO IMAGE

DESCRIPTION

TECHNICAL FIELD:

The present invention relates generally to the field of computer video image manipulation, and more particularly to an improved means and method for identifying objects of concern within a video image such that identification of the objects can be maintained even as the objects move within the image.

BACKGROUND ART:

Manipulation of digitized video images, both still pictures and moving video presentations, is an important aspect of the present trend toward the introduction of "multimedia" into many aspects of our lives, as well as in modern aspects of more traditional endeavors such as, for example, the creation of motion pictures. A copending U.S. Patent Application No. 08/146,964, having an inventor in common with this present invention, teaches a method for converting a conventional "moving picture" video into a computer/user interface means. In accomplishing the method of that previous invention, it is necessary to identify, within the video presentation, particular objects of concern. As discussed in the above referenced disclosure, such identification can be quite laborious, and it was anticipated that methods for transferring some of that labor from the human operator to the computer might be developed in the future. It was disclosed that the designation of "hot spots", consisting of objects within a moving video, were, "....accomplished by viewing each key frame and, at least until a more automated system is developed therefor, manually designating which, if any, objects or items of interest in the key frame are to be designated as the hot spots." (Reference numbers relating to the prior designation have been deleted in this quotation.) This present application is directed to a method and means for automating the identification of such objects and maintaining such identification through time. Although the present inventive method is intended to be used in conjunction with the production of interactive computer interface systems, it is not restricted to such applications.

An object in animated, or other specially prepared moving video images, can be rather easily identified, since such object can be created according to a specific easily distinguishable criterion (such as color, or the like) or, indeed, the pixel location of the object can be made a part of the data which describes the object within the computer even as the object is created. However, objects within a live action video, which video has not been specially produced nor specially prepared, cannot be so easily segregated.

Prior art methods for identifying objects in a video image, such that the object is defined according to computer understandable criteria, have included identifying edges, colors or color patterns and/or brightness levels which define the object. Such methods have been relatively effective for the identification and/or manipulation of still video images. For example, an object can be distinguished by known methods for automatically defining the outer edges thereof, and the object can then be operated upon. As examples of such operations, the object can be moved within the image, removed from the image, or changed in

color, luminosity, or the like. More in the context of the present invention, the object could even then, once the image is defined in terms of a bit map, be used in the manner of an icon or a "hot spot", such that clicking on the area of the image within the object could cause the computer to initiate a response or further interaction with the end user. It should be remembered, however, that this sort of procedure does not transfer well into the realm of moving video images. Firstly, keeping track of the location of objects within a moving video image by storing a bit map of all such objects for all frames of the moving image would require a morass of data which would tax a computer's data storage capacity and slow down the operation of the computer. Secondly, although the amount of user interaction and labor required to perform the above described operations is well tolerable when working with a single still video image, an attempt to repeat such an operation thirty or so times for each second of a moving video would quickly reveal that this method is outside the realm of practicality.

One accepted method for separating objects within a moving video image has been based upon the color of portions of the image. One skilled in the art will be familiar with the "blue screen" method wherein portions of an image which are of a specific color (often, but not necessarily, blue) can be selectively removed from an image. This technique was used in television prior to the advent of digital image manipulation, and has been found to work well also when applied to the field of digital image manipulation. While this method works well for its intended purpose, it will generally only successfully distinguish, for example, a background from the foreground object(s). Furthermore, it requires a special setting in that the object(s) of concern must be originally "shot" (meaning photographed, video taped, or the like) against the special background color. Most importantly, although the background is distinguished from the foreground objects such that a computer can calculate the location of the objects in order to perform operations thereon (such as overlaying the objects upon a different background), different objects are usually not sufficiently identifiable in terms usable by the computer such that the objects can serve as means for computer/user interaction. Moreover, even in those special situations in which a video scene can be shot in front of a blue background or the like, and even in those unusual instances wherein there may be only a single object in the foreground such that there will be no confusion between objects, such prior art solutions do not address the problem of extensive data storage requirements and drain on computation resources, as discussed above.

Methods for identification of edges or complex color patterns within a video image are more effective for segregating specific "real world" portions of a video image, as compared to the more artificially induced background "blue screen" methods. However, such edge or color identification methods generally require relatively sophisticated computer analysis, and so are not suitable for real time image tracking, at least unless a great deal of expensive computing power is dedicated to such tracking. Even where a practically unlimited quantity of computer resources are available to the task, attempting to track moving objects within a video image according to such complex criteria has proven to be undesirably complicated. Where more than one object is to be tracked within the video, or where the objects are rapidly moving and/or changing relative shape within the video, the problems associated with such methods are exacerbated.

It has been brought to the inventor's attention that several prominent manufacturers of computer

products have a need for a better means and/or method for identifying moving objects within video images such that the objects may be followed by a computer, in order to implement their own products. However, in spite of the fact that some of these companies have extensive research budgets and large and experienced research staffs, they have turned to the present inventor seeking a solution to this known problem.

To the inventor's knowledge, no workable method has existed in the prior art for quickly and easily identifying, for computer tracking and manipulation, objects within moving video images which is inexpensive and easy to implement and reliable. All prior art methods have either been extremely labor intensive and/or have required an inordinate amount of computing power to implement (or, even worse, have required an inordinate amount of computing power for an end user to utilize the product of such methods) and/or have not reliably identified objects such that a computer can track the objects within a video presentation without "losing" the objects or confusing them with the backgrounds or other objects in the video.

DISCLOSURE OF INVENTION:

Accordingly, it is an object of the present invention to provide a method for easily identifying objects within a digitized moving video image such that the objects can be tracked by a computer.

It is still another object of the present invention to provide a method for identifying objects within a digitized moving video image which requires a minimum of operator intervention and labor.

It is yet another object of the present invention to provide a method for identifying objects within a digitized moving video image which does not require a great deal of computer power to implement.

It is still another object of the present invention to provide a method for identifying objects within a digitized moving video image which identifies the objects in such a manner that a computer can store the identifying criteria using a minimum amount of data storage resources.

It is yet another object of the present invention to provide a method for identifying objects within a digitized moving video image which will allow a computer to easily follow the object as it moves through the image, without overburdening the computer's resources.

It is still another object of the present invention to provide a method for identifying objects within a digitized moving video image which will allow the objects to serve as gateways for computer/user interaction in the manner of computer icons, and the like.

Briefly, the preferred embodiment of the present invention is an automated method for identifying objects within a digitized moving video image such that a computer can readily identify the objects, without overburdening the computer, even as the objects change shape or move within the video image. The inventive moving object identification method first identifies an object within a single frame of a moving video image. The initial identification can be accomplished using extensive originating user input, or by more automated methods. In a first preferred embodiment of the invention, conventional automated software methods are used for initial object identification. In an equally preferred alternate embodiment of the invention, a unique combination of software and hardware components are used for the initial object

identification.

Although the initial object identification, according to the present invention, is accomplished either according to known methods or a new combination of known methods and means, a primary inventive aspect of the present invention is that initial identification is followed by a conversion operation wherein only such data as is necessary to reliably track the object is retained. Therefore, very little recomputing will be required to track the object as the object moves and/or changes shape from frame to frame in the moving video presentation.

Regarding this present invention, it should be understood that computers will play a part both in the inventive method for identifying objects within a video presentation and, also, computers (more than likely, different computers than those used for originally identifying the objects) will be employed to eventually use the end product of the present invention. In this regard, there will be reference hereinafter to "originating user(s)", those being the users who use the present inventive method for creating a computer/user interface, or other appropriate application of the inventive method as will be disclosed herein. References to "end user(s)" will be to those persons who, rather than directly using the present inventive method, will use the computer/user interface means produced by the originating user(s) according to the present inventive method.

An advantage of the present invention is that it does not require an expensive powerful computer to be implemented.

A further advantage of the present invention is that it does not require an expensive powerful computer in order to use the end product of the method.

Still another advantage of the present invention is that it does not require a great deal of user input.

Yet another advantage of the present invention is that it allows a computer to reliably track objects within a digitized moving video, clearly distinguishing the objects from background and other objects such that the objects of interest can be acted upon by a computer and/or by a computer user acting in conjunction with a computer.

These and other objects and advantages of the present invention will become clear to those skilled in the art in view of the description of the best presently known mode of carrying out the invention and the industrial applicability of the preferred embodiments as described herein and as illustrated in the several figures of the drawing.

BRIEF DESCRIPTION OF DRAWINGS:

Fig. 1 is a flow chart depicting a moving object identification method according to the present invention;

Fig. 2 is a depiction of a video image such as is acted upon according to the present inventive method;

Fig. 3 is a detailed view of the an object of interest of **Fig. 2**, showing the relationship of the object to a hot spot;

Fig. 4 is a detailed view of a hot spot defining tag according to the present invention;

Fig. 5 is a more detailed flow diagram of the tag objects operation of **Fig. 1**;

Fig. 6 is a more detailed flow diagram showing an alternative sequence of substeps of the tag objects operation of **Fig. 1**; and

Fig. 7 is a diagrammatic representation showing the relationship of a simplified object to a camera.

BEST MODE OF CARRYING OUT THE INVENTION:

The best presently known mode for carrying out the invention is a moving object identification method instituted, primarily, through the use of a computer. The predominant expected usage of the inventive moving object identification method is in the making of interactive computer/user interface systems, although the method might be used in essentially any application wherein it is necessary to track an object within a moving video image.

The inventive moving object identification method is depicted in a flow chart in **Fig. 1**, and is designated therein by the general reference character 10. A 'criteria choice' operation is the first step of the moving object identification method.

Fig. 2 is a diagrammatic representation of a video image 14 being displayed on a display screen 16 of a computer 18. The video image 14 is a single "frame" of a moving video presentation which has been digitized according to conventional methods such that the video image 14 can be stored in the computer 18 and displayed thereby on the display screen 16. It should be noted that, although one skilled in the art will recognize that there are substantial differences in format between a motion picture film, a television signal and a digitized video presentation, much of the terminology relating to these different media is transferable among the three. Therefore, although the video image 14 of **Fig. 1** is an element of a digitized moving video presentation, it is convenient to discuss the video image 14 as though it were a "frame" of a photographically encoded moving picture presentation. Those skilled in the relevant art will readily understand the relationship between the digitized video image 14 and conceptually corresponding moving picture "frames". This method of reference is frequently used in the art, and the meaning of a "frame" of video will be recognized as corresponding to the single video image 14, although no actual physical embodiment of a frame may exist.

In the view of **Fig. 2**, it can be seen that an object 20 is depicted upon a background 22. The background 22, in this sense, is all of the video image 14 which is not the object 20 of concern. The background 22 may include nondescript and generally undefinable portions and, as in the example of **Fig. 2**, other item(s) 24, which other item(s) 24 are potentially definable portions of the background 22 that might, or might not, be separately defined as additional objects of interest.

Methods for identifying the object 20 within the video image 14 such that the computer 18 can perform some operation thereon, such as changing the color of the object 20 or deleting or moving it within the video image 14, or the like, are known in the art. In the best presently known embodiment 10 of the present invention, a well known and widely practiced edge detection technique known as *Sobel's Algorithm* (discussed in detail in *Digital Image Processing* by *Gonzolez* and *Woods*, Addison-Wesley Publishing, 1992, p. 197-201 and 416-420) is used to detect an edge 26 of the object 20. The edge 26

is, according to *Gonzolez* and *Woods*, a boundary between two regions of an image (the two regions being the object 20 and the background 22, in the present example of **Fig. 2**) which two regions have relatively distinct luminosity characteristics, although it is certainly conceivable that the edge 26 be defined according to color characteristics, or the like, in addition to or instead of by the luminosity characteristics.

A plurality (six, in the example of **Fig. 2**) of tags 28 are depicted in **Fig. 2**. The combination of one or more of the tag(s) 28 on the object 20 define a hot spot 30. That is, the hot spot 30 and the object 20 are perceptually, to an end user, essentially generally one and the same. However, as defined herein, the object 20 is merely an area on the video image 14 which can be perceived as being a definable entity by a human observer, while the hot spot 30 is defined such that the computer 18 can recognize and track the hot spot 30, even as the hot spot 30 changes position and/or shape within the video image 14.

Fig. 3 is a more detailed diagrammatic view of the object 20 of **Fig. 2**. In the view of **Fig. 3** it can be seen that the hot spot 30 is defined by the tags 28 such that a hot spot border 32 is established. It can ben seen, then, that the hot spot 30 and the object 20 occupy generally, but not exactly, the same portion of the video image 14 **(Fig. 1)**. This will have no practical significance in the primary intended applications, since any undefined areas 34 which are within the object 20 confines and yet without the limits of the hot spot 30 will be at the outer limits of the object 20. An end user attempting to "click on" the object will naturally tend to click near the center of the object 20 and, therefore, also near the center of the hot spot 30. An end user will instinctively not expect a response from a click in the undefined areas 34 which lie near the edges of the object 20, even though the end user will not see nor have a way of knowing the exact locations of the hot spot border 32. As can be appreciated, the greater the quantity of properly distributed tags 28, the greater will be the correlation between the hot spot 30 and the object 20.

Fig. 4 is a detailed diagram of a representative tag 28a of **Fig. 2**. The edge 26, as previously discussed herein, is defined according to principles and methods well understood in the art. In particular, in the example of the best presently known embodiment 10 of the present invention, the widely practiced *Sobel's Algorithm* is used to identify the edge 26. It should be noted that the process of edge identification will generally identify edges promiscuously, such that not only the edge 26 edge of the object 20, but also any other distinct edges (not specifically identified) which might exist in the video image 14 **(Fig. 2)** will be so identified.

Referring again to the flow chart of **Fig. 1**, it can be seen that the 'criteria choice' operation 12, as described herein in relation to the example of the best presently known embodiment 10 of the present invention, is the choice to use *Sobel's Algorithm* to detect edges (including the edge 26 of particular interest) according to luminance criteria. It may be, in some applications of the moving object identification method, that the 'criteria choice' operation 12 is fixed and unchanging for all uses of that application, such that the originating user will not accomplish the 'criteria choice' operation 12 each time the inventive moving object identification method 10 is accomplished.

Referring again to the flow diagram of **Fig. 1**, it can be seen that a 'tag objects' operation 36 is accomplished following the 'criteria choice' operation 12. As will be described in more detail hereinafter, in the best presently known embodiment 10 of the present invention, originating user intervention is required to accomplish the 'tag objects' operation 36, although it is within the scope of the present

invention that some means or method might be developed in the future to essentially completely automate the 'tag objects' operation 36, or at least to further assist the originating user in order to minimize the need for user intervention.

In the best presently known embodiment 10 of the present invention, in the 'tag objects' operation 36 the user places the tags 28 around the borders of the object 26 of interest by clicking on the edge 26 of the object. **Fig. 6** is a more detailed flow diagram of the 'tag objects' operation 36 of **Fig. 1**. In the view of **Fig. 5**, it can be seen that an 'affix tag to object' substep 38 precedes a 'relate tag to edge' substep 40. According to the best presently known embodiment 10 of the present invention, the originating user will "click on" the edge 26 of the object 20, as by using a mouse 40 (**Fig. 1**) to move a cursor 42 to the edge 26 whereafter the originating user will then depress a button 44 of the mouse 40. It should be remembered that the originating user, at this point in the process, will have to visually perceive the presence of the edge 26.

Referring again to the view of **Fig. 4**, a plurality of pixels 44 are identified. One skilled in the art will recognize that the display screen 16 has a plurality of the pixels 44 thereon, the exact quantity of which will be a function of the characteristics of the display screen 16 itself, as well as of the hardware and software controlling the video image 14. In order to conveniently describe the present inventive moving object identification method 10, the pixels 44 are illustrated on a grid such that the pixels 44 appear to be generally square, although one skilled in the art will recognize that actual physical pixels (not shown) may be round, rectangular, or other shape.

In the present example, if the originating user clicks on a selection point 46 (the selection point 46 being established when the originating user places the cursor 42 (**Fig. 2**) and depresses the mouse button 44), the tag 28a is placed with the center thereof being the selection point 46. In the example of **Fig. 4**, the tag 28a is 9x9 pixels 36 in size, the implications of which size will be discussed in more detail hereinafter.

Referring again to **Fig. 5**, the 'affix tag to object' operation 38 described above, is followed by a 'relate tag to edge' operation 48. According to the best presently known embodiment 10 of the present invention, in the 'relate tag to edge' operation 48, the edge 26 is automatically detected within the tag 28a, and an edge mask 50 is established on the edge 26 and at the center of the tag 28a. As previously discussed herein, for the purpose of illustrating the best presently known embodiment 10 of the present invention, *Sobel's Algorithm* was chosen as the means of edge detection in the 'criteria choice' operation 12 (**Fig. 1**). Those familiar with *Sobel's Algorithm* will note that a 3x3 mask is customarily used in that edge detection process, and so it is convenient, when using *Sobel's Algorithm*, for the edge mask 50 to be three pixels 44 by three pixels 44 in size. It is also convenient to have the tag 28a be n by n pixels 44 in size where n is an integer greater than 1, as is illustrated by the 9x9 sized tag 28a of **Fig. 4**. Of course, a modified Sobel's Algorithm could be used such that the edge mask 50 would be of another size or, alternatively, an entirely different edge detection method might be employed whereby the edge mask 50 could be of essentially any size practical according to the other parameters of the present invention detailed herein.

It should be noted that, in the example of **Fig. 4**, as previously discussed herein, the originating

user has precisely clicked on the edge 26 such that the selection point 46 falls on the edge 26. In actual practice, it might be that the originating user's estimation of the location of the edge 2 might be less than precise such that the initial location of the selection point 46 would not fall precisely on the edge 46. This situation is rather easily corrected, as follows: When the tag 28a is placed in the 'affix tag to object' substep 38 and, thereafter, the edge 26 is detected and the edge mask 50 is centered on the edge 26 in the 'relate tag to edge' substep 48, if the edge mask 50 is not centered in the tag 28a, then the tag 28a is automatically repositioned by the computer 18 such that the edge mask 50 is centered on the tag 28a. It is contemplated that alternative methods might be employed for dealing with situations wherein there is no portion of the edge 26 detected within the tag 28a. For example, the originating user could be advised of a "no edge detected" condition, and told to try again. Alternatively, the search for the edge could be expanded a predetermined number of pixels 40 beyond the tag 28a. If the edge 26 is detected in such an expanded search then the position of the edge mask 50 and the tag 28a could be automatically adjusted accordingly. If, even after such an expanded search, the edge 26 is still not detected, then the originating user could be advised to repeat the 'affix tag to object' operation 38.

Fig. 6 is a flow diagram of an alternate 'tag objects' operation 36a which is not presently proposed as being the order of operations of the best presently known embodiment 10 of the present invention, but which, instead, illustrates a variant of the best presently known embodiment 10 which might be useful in some unforeseen circumstances. As can be seen from the flow diagram of **Fig. 6**, an alternate 'affix tag to object' substep 38a and an alternate 'relate tag to edge' substep 48a are reversed in order as compared to the correspondingly named substeps 38 and 48 of **Fig. 5**. This reversal of order requires some modification of the 'affix tag to object' substep 38 and the 'tag to edge' substep 48 which have been previously discussed herein.

As previously described herein, in the 'relate tag to edge' operation 48, edge detection (according to *Sobel's Algorithm* in the present example) occurs generally within the tag 28a. In the alternative example of **Fig. 6**, upon completion of the 'criteria choice' operation 12, the edge detection process is run on the entire video image 14 **(Fig. 1)** such that the edge 26 (as well as any other edges within the video image 14 which are not specifically delineated herein) is highlighted so that the originating user then can see the edge 26. By comparison, in the previously described 'affix tag to object' operation 38, the originating user could estimate where the computer 18 might detect the edge 26, whereas in this present alternate example, at the outset of the alternate 'relate tag to edge' substep 48a and precise location of the edge 26, as determined by the computer 18, is made known to the originating user such that the originating user can place the tag 28a precisely on the edge 26. After the originating user places the tag 28a on the edge 26 in the alternate 'relate tag to edge' operation 48a, then it is a relatively simple matter for the computer 18 to center the tag 28a, in the alternate 'affix tag to object' substep 38a, such that the edge mask 50 (which, according to the alternate 'tag objects' operation 36a, will already coincide with a portion of the edge 26) is symmetrically arrayed about the edge mask 50.

Following the 'tag objects' operation 36 (or following the alternative ' tag objects' operation 36a, as discussed above) is a 'select region of interest' ("select ROI") operation 52. In the 'select ROI' operation 52, the originating user will merely click on the side of the edge 26 wherein is the object 20 to designate

a region of interest ("ROI") 54, the ROI 54 being that portion of the video image 14 wherein the tag 28a and the object 20 coincide.

Following the 'select ROI' operation 52, the originating user is given a choice as to whether it is desired to place more tags 28 on the object 20 in a 'place more tags?' decision operation 56. If it is desired to place more tags 28, then the inventive process returns to the 'tag objects' operation 36. If the response to the 'place more tags?' decision operation 56 is no ("N"), then the process continues to a 'track objects' operation 28.

It should be noted that the operations of the inventive moving object identification method 10 which have been presented herebefore are accomplished on a single "frame" of the moving video image 14 (**Fig. 2**). Beginning now a discussion of how the inventive procedure continues as the video image 14 is changes through time, one skilled in the art will recognize that, in a digitized live action video production, the frame-to-frame attributes such as the size of the edge 26, color values of each of the pixels 44, luminosity characteristics, and the like, can each and all change due to various noise factors, changes in lighting, movements of the object 20 and other item(s) 24 (if present) in the video image 14 and/or movements of a camera (not shown) which originally captured the video image 14 in relation to the object 20, and the like.

Preliminary to discussing the tracking of the object 20 through time by means of the tags 28, it is anticipated by the inventor that certain parameters will be used for determining if the tags 28 continue to accurately represent the object 20 such that the hot spot 30 and the object 20 generally coincide. According to the anticipated best presently known embodiment 10 of the present invention, such parameters are: Luminosity, color and tag sensitivity.

Regarding luminosity, as part of the 'track objects' operation 58, the computer 18 will calculate the average luminosity within the tag 28 on either side of the edge 26. The originating user may specify an amount, expressed as a percentage of luminosity difference, by which this value can change before each respective tag 28 is flagged by the computer 18 as lost. (The fact and consequences of "tag lost" conditions will be discussed in greater detail hereinafter.)

Regarding color, the average value of the color of the pixels 44 within the ROI 54 is tracked by the computer 18 on a component-by-component (i.e. RGB) basis. The color parameter allows the originating user to specify the amount, expressed as a percentage, by which such average values can change before each respective tag 28 is flagged as lost.

Regarding tag sensitivity: This parameter addresses a phenomenon which is peculiar to the tracking of the three dimensional object 20 in two dimensional space (the video image 14), which describes the native environment of the inventive moving object identification method 10. **Fig. 7** is a diagrammatic view illustrating a simplified (rectangular) object 20a shown in relation to a camera 60. A vertical edge 26a (chosen arbitrarily for purposes of this example only, since a horizontal edge (not specifically delineated) or other such edge could also be used to illustrate the point being made) is highlighted in the view of **Fig. 7** for the purpose of this present discussion only. In relation to the view of **Fig. 7**, it can be seen that the simplified object 20a might rotate according to a first rotational arrow 62 such that the perceived length of the vertical edge 26a will remain generally constant as viewed from the camera 60. Alternatively, the

simplified object 26a might rotate according to a second rotational arrow 64 such that the perceived length of the vertical edge 26a will vary according to the amount of rotation. Indeed, should the simplified object 20a rotate a full 90 degrees in either direction indicated by the second rotational arrow 64, the vertical edge 26a, as viewed from the perspective of the camera 60, will become essentially infinitesimally short. It can be appreciated that edges 26 which are not perfectly vertical or horizontal will experience this phenomenon in varying degrees depending upon the angle of such edge 26 relative to the camera 60 and further depending upon the actual direction of rotation of the object 20 (**Fig. 2**) as illustrated by the simplified object 20a (**Fig. 7**). Also, it can be appreciated ,in light of the above discussion, that movement of the camera 60 relative to the simplified object 20a will produce generally the same sort of relative lengthening and shortening of the edge 26 (as typified by the vertical edge 26a) as will movement of the simplified object 20a relative to the camera 60.

Referring now again to the view of **Fig. 4**, a sensitivity zone 66 is that portion of the tag 28a which is not also within a subtag 68. In the example of **Fig. 4**, the subtag 68 is 5x5 pixels 44 in size. The subtag 68 is centered within the tag 28a. In practice, the subtag 68 might, or might not, be the same size as, and thus completely coincident with, the edge mask 50. The vertical edge 26a is depicted in the example of **Fig. 4** as being horizontally centered within the representative tag 28a (as it will be according to either the 'tag objects' operation 36 or the alternative 'tag objects' operation 36a as previously discussed herein. As can be appreciated from the view of **Fig. 2**, as the simplified object 20a rotates as indicated by the second rotational arrow such that the vertical edge 26a moves away from the vertical, the apparent length of the vertical edge 26a from the perspective of the camera 60 will become increasingly shorter. When the vertical edge 26a becomes sufficiently short that it is within the subtag 68 but not within the sensitivity zone 66, then a tag lost condition will be indicated for that particular tag 28. As can be appreciated in light of the above discussion, the "sensitivity" for responding to this condition can be adjusted by adjusting the size of the sensitivity zone 66. That is, the larger is the subtag 68 in relation to the tag 28, then the smaller will be the sensitivity zone 66 and the quicker will a tag lost condition be indicated as loss of the edge 26 (as illustrated by the example of the vertical edge 26a) occurs. (It will be noted that, referring to the view of **Fig. 7**, the representative tag 28a is not actually superimposed on the simplified object 26a such that the representative tag will be visible thereon. Rather, the representative tag 28a is shown in the view of **Fig. 7** simply for the purpose of denoting that portion of the vertical edge 26a that is also shown in the view of **Fig. 4**.)

Fig. 8 illustrates an equally preferred alternate embodiment 810 of the inventive moving object identification method. In the view of **Fig. 8**, it can be seen that a complex object 20b has distributed thereon a plurality of physical tags 28b. The physical tags 28b are transponders which are capable of relaying their relative position to the camera 60 such that such relative position(s) can be recorded along with the video image 14 (**Fig. 2**) thereof. Such transponders as constitute the physical tags are now commonly available. These operate according to any of several technologies, including radio triangulation and ultrasonic triangulation principles. For example, an ultrasonic system which is capable of very precise location information in three dimensions, as well as attitudinal information (roll, pitch and yaw) of the physical tags 28b, while not creating audible noise and while being generally impervious to interference

from ambient noises, is marketed as a developers kit by Logitechtm. It should be noted that it is known in the field to attach a single sensor to objects (such as people) for recording the position of such objects relative to a camera during the making of a motion picture. However such objects as previously known and used are not sufficiently sensitive, nor do they need to be when used for the purposes for which they have been employed in the prior art, to distinguish the limits of such objects.

When the physical tags 28b are implemented as illustrated in **Fig. 8**, it is a simple matter to transfer the recorded location of the physical tags into software such that the computer 18 will be able to define the object(s) 20 according to (software) tags 28, as described herein. After the locations of the physical tags 28b are transferred from the original recording media such that the computer 18 will have a record of such locations for the video image 14, then the location of the physical tags 28b is not significantly different from the location of the tags 28 that were originally created in software as previously discussed herein.

It should be noted that, whether the (software) tags 28 or the physical tags 28a are initially used, there will occur some general conditions not specifically addressed previously herein. For example, it will often occur that as the object(s) 20 rotate or move within the video image 14 that the edges 26 will appear to cross over when view from the two dimensional perspective. (That is, for example, when an object 20 turns completely around such that the right becomes the left and the left becomes the right as view from a stationary perspective, then the edges will appear to cross over at some point in between.) Actually such variables pose little or no problem in relation to the present inventive moving object identification method 10. Where a tag 28 crosses over the hot spot border 32 such that the tag 28 is inside the hot spot border 32, it can simply be temporarily ignored until it again crosses over the existing hot spot border 32 and emerges from within the hot spot 30.

Fig. 9 illustrates the relationship of the computer 18 which is used to assist in performing the present inventive method to an end user computer 18a upon which an end product such as an interactive computer/user interface 70 will be used. The interactive computer/user interface 70 is shown in the view of **Fig. 2** embodied in a CD ROM. Of course, the interactive computer/user interface 70 will generally only constitute a relatively small portion of the content of a typical CD ROM in which it might be included. Having been produced on the computer 28, the interactive computer/user interface 70 will be incorporated into the CD ROM (or, alternatively, into a floppy disk or other means of communicating to the end user computer 18a. Then that CD ROM will generally be copied and that CD ROM (or, more generally, a copy thereof) will be transferred to the end user computer 18a where the interactive computer/user interface 70 will be introduced thereinto and then displayed upon an end user display screen 16a of the end user computer 18a.

Returning again to a consideration of the "tag lost" condition referenced several times previously herein, it will be noted that a single lost tag 28 will generally not be fatal to the continuation of the inventive moving object identification method 10. As noted herein, the location of the tags 28 for defining the hot spots 30 is associated with the video image 14 by the originating user according to the present inventive moving object identification method 10, with the aid of the computer 18. As previously discussed herein, the originating user, will be able to seen the tags 28 and the hot spot borders 32, when the end user will

not and will, therefore, generally not be able to distinguish between the object 20 and the corresponding hot spot 30. The tag 28 can become "lost" when an edge 26 is foreshortened, when the tag 28 moves so quickly that it is cannot be followed according to the present inventive method as presently embodied, or when "noise" of one sort or another masks the tag 28. When the originating user is advised of a tag lost condition, the originating user will be able to make an assessment as to whether or not there remains sufficient correlation between the hot spot 30 and the object 20. Should these still be appear to be essentially the same, then the 'track objects' operation 58 will be allowed to continue tracking the object 20 and storing associated locations of the tags 28 therefor. Should a determination be made that there is no longer sufficient correlation between the hot spot 30 and the object 20, then the originating user can restart the inventive moving object identification method 10 beginning at the chronological point in the run time of the video image 14 wherein it is determined that such is required.

It should be noted that modifications to the present inventive method are contemplated which will minimize the computational time required in tracking the tags 28 in the computer 18 of the originating user. Of course, this will not be a consideration regarding the end user computer 18a, since the location of the tags 28 has been recorded by the originating computer 18 and the end user computer 18a will simply note that proper location of the tags 28 relative to the video image 14. However, during the actual accomplishment of the moving object identification method 10, there can be a significant amount of time involved in the track objects operation 58, which operation is performed by the computer 18. In order to minimize the amount of computational time required to track the tags 28, the following methods have been found to be of practical value: It is often convenient to assume that a tag 28 has not been moved from its previous location. This assumption can be quickly checked and, if it is valid, then the time expended in recomputing a tag 28 location is saved. When the previous assumption proves to be incorrect, it is further convenient to assume that a previous direction of movement remains constant. Therefore, the computer 18 can try placing the tag 28 in a location determined by the previous direction and amount of displacement of that tag 18, thereby avoiding having to move the edge mask 50 over a larger area in search of the edge 26. Of course, should all such assumptions prove to be incorrect, the computer 18 will have to expand its area of search or else report a tag lost condition as previously discussed herein. It is expected that these and other refinements of the present inventive moving object identification method 10 will be developed and improved as the inventive method is implemented and adapted to various applications.

Various modifications may be made to the invention without altering its value or scope. For example, although the best presently known embodiment 10 of the present inventive method uses edge identification to originally define the object 28 as an aid to placing the software tags 28, means such as color identification could be equally well applied.

Another likely modification would be to further automate the location of the tags 28. For example, where the 'alternative tag objects' operation 36a is employed such that the edge 26 is identified before the tag 28 is placed, software might be written wherein the originating user could merely click within the object and a plurality of rays could be generated from the location thus selected outward with the intersection of the rays and the edge 26 being marked as tags 28. Further refinements on this method might allow the

originating user to select from among the tags 28 thus created.

All of the above are only some of the examples of available embodiments of the present invention. Those skilled in the art will readily observe that numerous other modifications and alterations may be made without departing from the spirit and scope of the invention. Accordingly, the above disclosure is not intended as limiting and the appended claims are to be interpreted as encompassing the entire scope of the invention.

INDUSTRIAL APPLICABILITY:

The inventive moving object identification method 10 is intended to be widely used in the multimedia industry. The predominant current usages are for the identification and tracking of objects within a moving video image such that the objects can be used as though the object were a graphical icon or a similar means for interaction between an end user and a computer.

In practice, the inventive moving object identification method 10 will be incorporated into many more extensive methods wherein the identification of moving objects 20 within a video image 14 is a necessary or desirable means for interacting with the product of such more extensive method. A principle example will be in the production of the interactive computer/user interface wherein the end user may click on the object 20 to access further information or initiate some other action. For instance, where further information is available about such object 20 the end user can access the further information by clicking on the object.

According to the present inventive method, there can be hot spots 30 within hot spots 30 on a given screen and/or clicking upon one hot spot can bring up another screen or a partial screen which subsequent screen(s) also contain other hot spots. Thereby, the degree of interactivity of a program or interface is greatly enhanced over the prior art.

Since the moving object identification method 10 of the present invention may be readily integrated into existing and future end product creation methods, it is expected that it will be acceptable in the industry as a new and useful tool available to those engaged in the creation of computer software and software/hardware combination products. For these and other reasons, it is expected that the utility and industrial applicability of the invention will be both significant in scope and long-lasting in duration.

CLAIMS

1. A method for distinguishing an object within a digitized moving video image, comprising:

a) placing a quantity of tags on the moving video image such that said tags generally outline the object;

b) causing said tags to move as the moving video image advances, such that said tags continue to generally outline the object even as the object moves and changes shape in the moving video image.

2. The method of claim 1, wherein:

a) said tags comprise a two dimensional array of pixels; and

b) said tags are initially placed on the moving video image such that said tags include an edge of the object.

3. The method of claim 1, wherein:

a) said tags are initially placed on the objects by means including a plurality of transponders, such that the transponder defines the edge of the object.

4. The method of claim 1, wherein:

a) the location of said tags is stored in a computer such that a hot spot area is identifiable by the computer, the hot spot area being that portion of the video image defined by said tags.

5. The method of claim 1, wherein

a) each of said tags is placed on the object when an originating user clicks on a desired location for each of said tags.

6. The method of claim 1, wherein:

a) each of said tags includes a smaller subtag therein; and

b) an edge within at least one of the tags is reported as lost when that edge falls within the subtag and further when that edge does not fall within a sensitivity zone, the sensitivity zone being that portion of the tag which is not within the subtag.

7. A method for creating a hot spot on a display screen of a computer, comprising:

a) locating a plurality of tags such that said tags generally define the shape of an object in a digitized moving video image;

b) causing said tags to maintain their relative position on the object as the object

moves in the moving video image; and

c) recording the location of said tags such that a portion of the moving video image bound by said tags is defined as the hot spot.

8. The method of claim 8, wherein:

a) at least one of said tags is located on an edge of the object.

9. The method of claim 8, wherein:

a) the tag is initially placed on the object such that the edge falls within the tag.

10. The method of claim 8, wherein:

a) as the edge moves in the moving video image the tag is moved such that the edge remains centered within the tag.

11. The method of claim 8, wherein:

a) a region of interest is defined within the tag such that the region of interest is that portion of the moving video image wherein the tag and the object coincide; and

b) the tag is reported as lost when the color of the region of interest changes according to a preset limit.

12. The method of claim 8, wherein:

a) a region of interest is defined within the tag such that the region of interest is that portion of the moving video image wherein the tag and the object coincide; and

b) the tag is reported as lost when the luminosity of the region of interest changes according to a preset limit.

13. The method of claim 8, wherein:

a) when the edge becomes foreshortened within the tag such that the edge no longer extends into a predefined sensitivity zone then the tag is lost.

14. The computer/user interface of claim 13, wherein:

a) the sensitivity zone is that portion of the moving video image which falls within the tag and without a subtag where the subtag is a smaller fully enclosed subset of the tag.

15. The method of claim 14, wherein:

a) the sensitivity zone is at least as large as an edge detection mask, the edge detection mask being an array of pixels used for detecting the edge.

16. The method of claim 7, wherein:

a) the tags are visible to an originating user as the method for creating a hot spot on a display screen of a computer is accomplished; and

b) the tags are not visible to an end user such that the object will be generally indistinguishable from the computer identifiable hot spot by the end user.

17. A method for tracking an object within a moving video image, comprising:

a) choosing a criteria by which the object can be initially defined;

b) placing a tag on the object according to the criteria chosen in step a;

c) selecting a region of interest within the tag, the region of interest being that portion of the tag which coincides with the object; and

d) repeating steps b and c until a hot spot defined by a plurality of the tags generally coincides with the object.

18. The method of claim 17, and further including:

a) following step d, causing each of the plurality of tags to track the object according to the criteria chosen in step a.

19. The method of claim 17, wherein:

a) an originating user can restart the inventive method when a hot spot border no longer generally defines the object, the hot spot border being a line interconnecting the hot spots which line is made visible to the originating user superimposed on the moving video image.

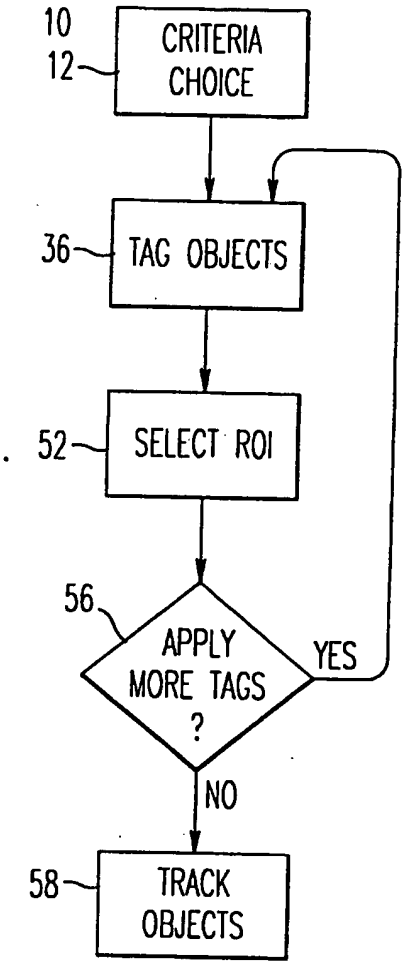

10
12
CRITERIA CHOICE
36
TAG OBJECTS
52
SELECT ROI
56
APPLY MORE TAGS ?
YES
NO
58
TRACK OBJECTS

FIG. 1

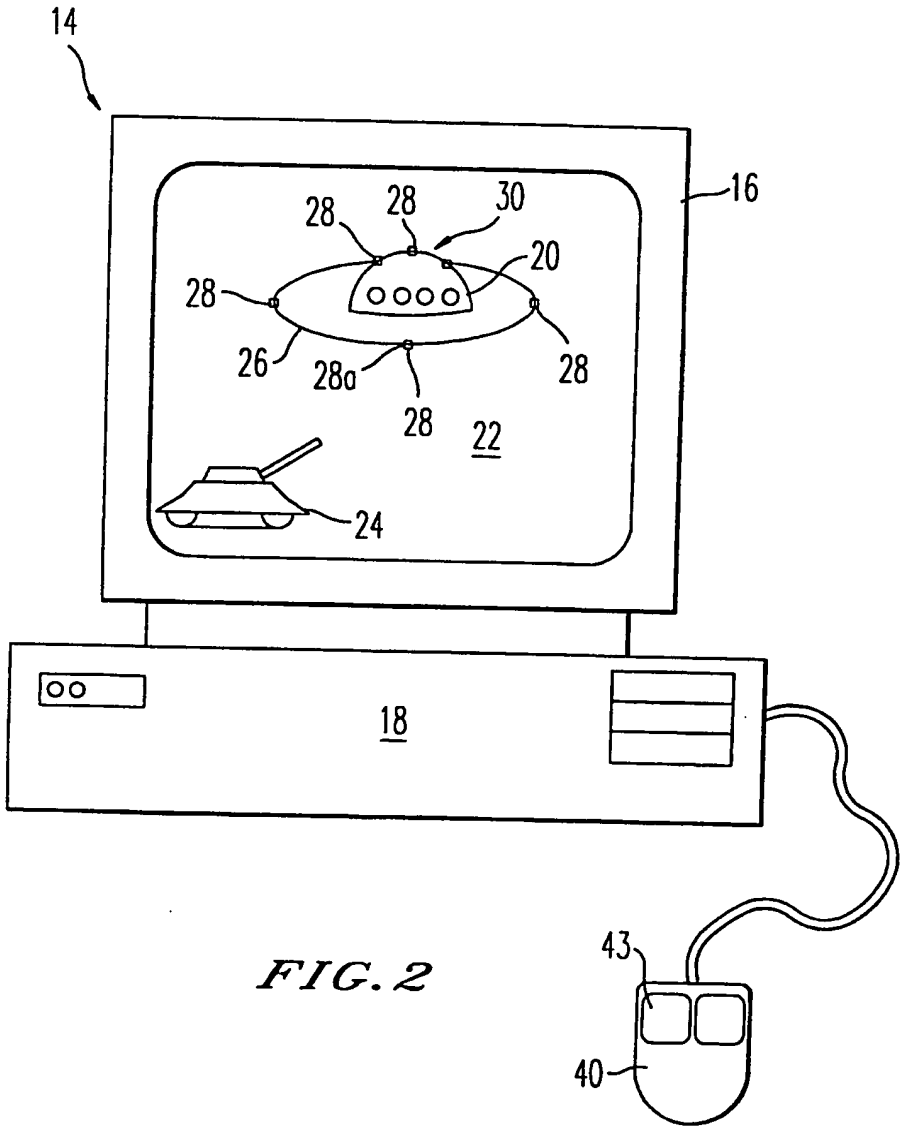


FIG. 2

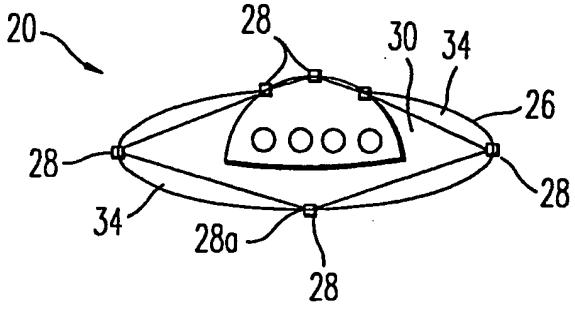


FIG. 3

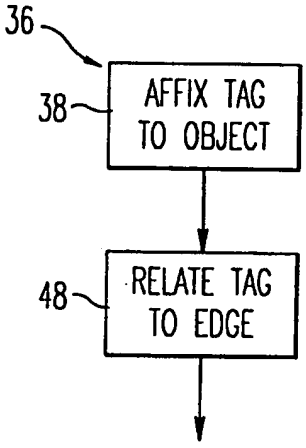


FIG. 5

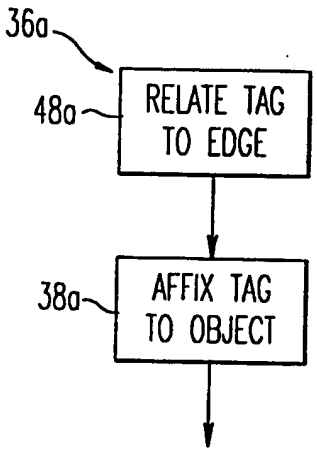


FIG. 6

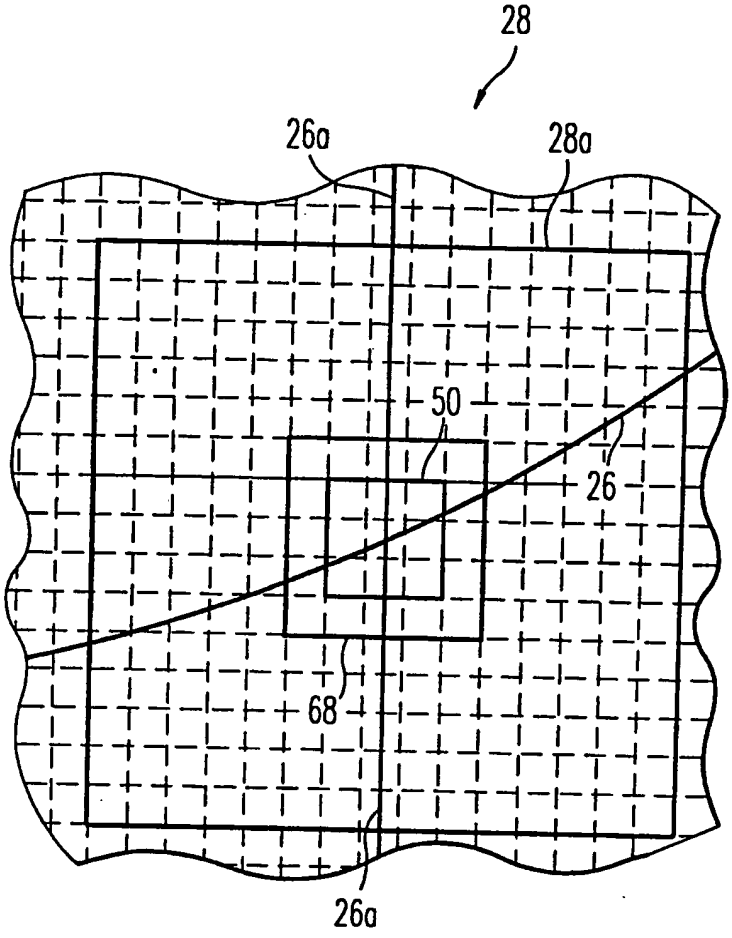


FIG. 4

FIG. 7

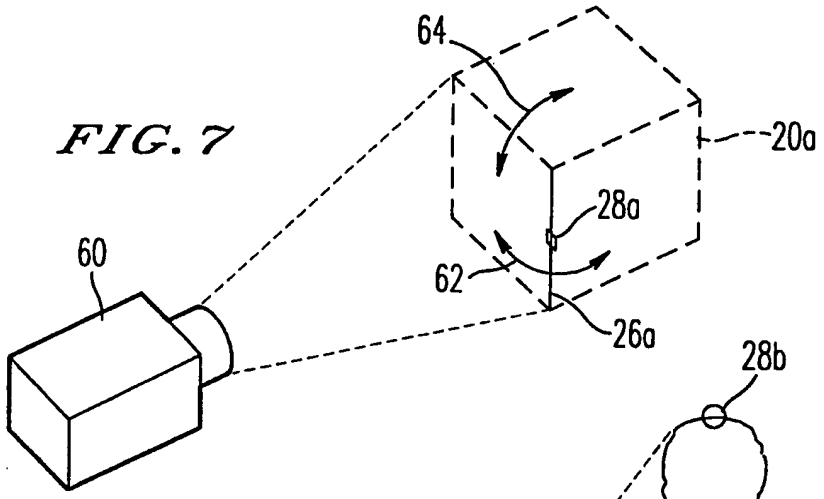


FIG. 8

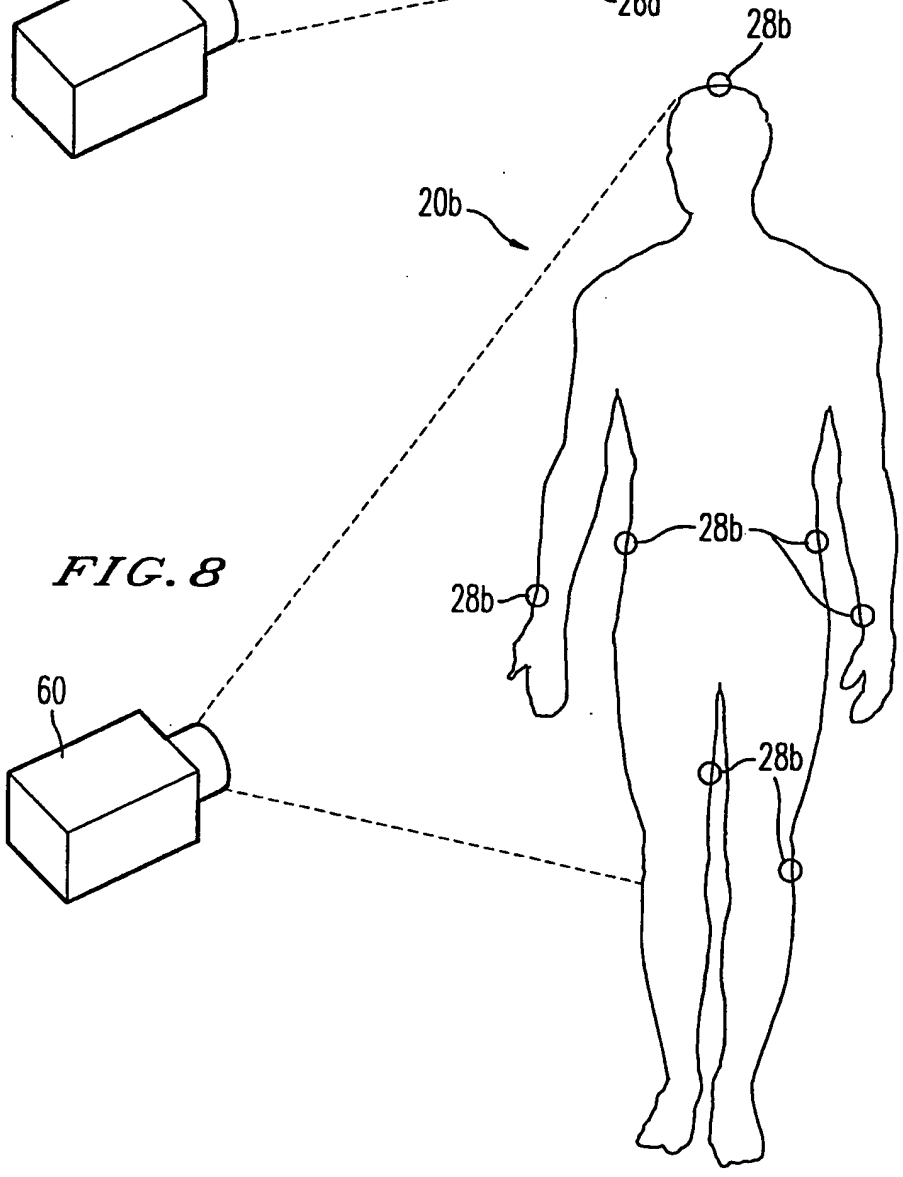
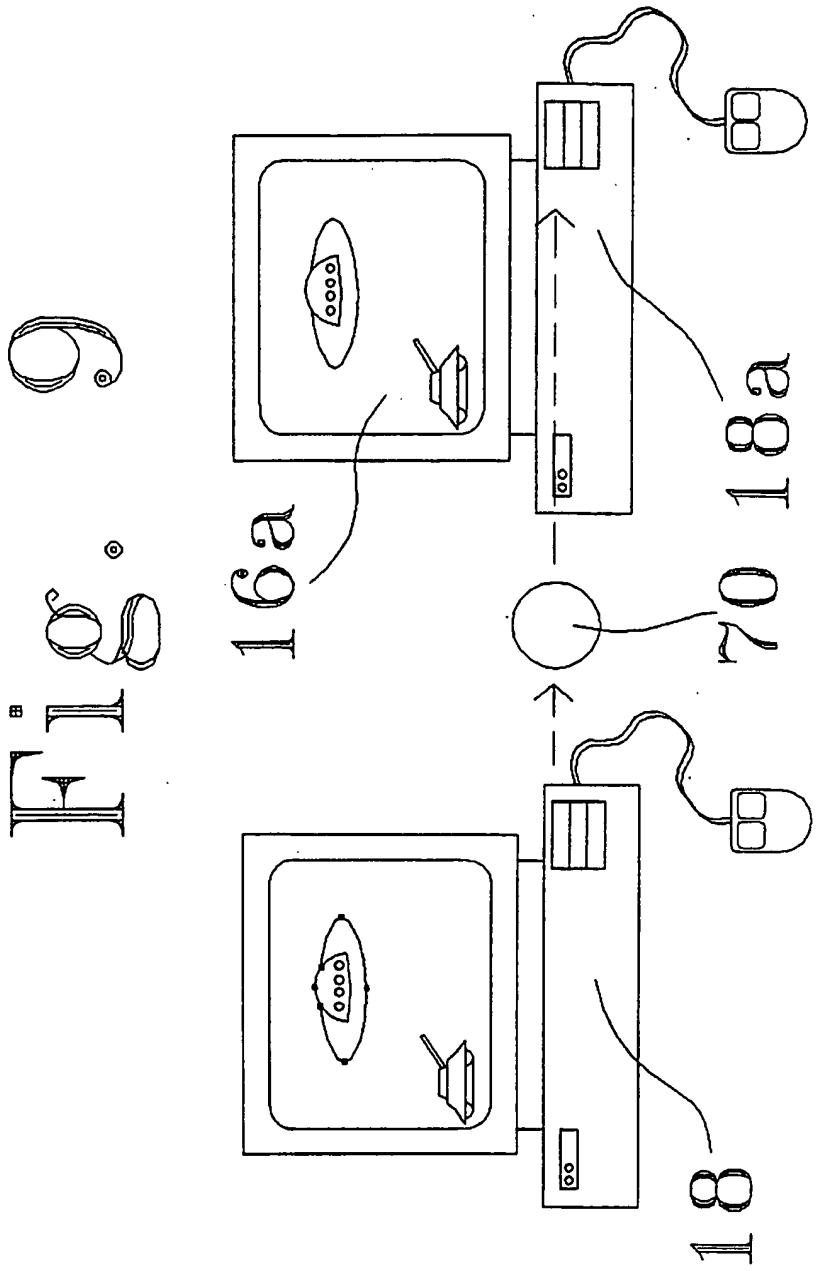


Fig. 6



INTERNATIONAL SEARCH REPORT

International application No.
PCT/US96/06542

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :G06K 9/00
US CL :382/107
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 382/107, 106, 236, 238; 348/409, 412, 413, 414, 416

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US, A, 5,243,418 (KUNO ET AL) 07 SEPTEMBER 1993, ENTIRE DOCUMENT.	1-19
Y	US, A, 5,311,305 (MAHADEVAN ET AL) 10 MAY 1994, ENTIRE DOCUMENT.	1-19
Y	US, A, 5,329,368 (PLOTKE) 12 JULY 1994, ENTIRE DOCUMENT.	1-19
Y	US, A, 5,323,470 (KARA ET AL) 21 JUNE 1994, ENTIRE DOCUMENT.	1-19
Y	US, A, 5,177,794 (ABE ET AL) 05 JANUARY 1993, ENTIRE DOCUMENT.	1-19
Y	US, A, 5,111,511 (ISHII ET AL) 05 MAY 1992, ENTIRE DOCUMENT.	1-19

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be part of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search 08 AUGUST 1996	Date of mailing of the international search report 04 SEP 1996
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer for MICHAEL RAZAVI Telephone No. (703) 305-4713

Form PCT/ISA/210 (second sheet)(July 1992)*

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US96/06542

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y,E	US, A, 5,537,155 (O'CONNELL ET AL) 16 JULY 1996, ENTIRE DOCUMENT.	1-19